

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**



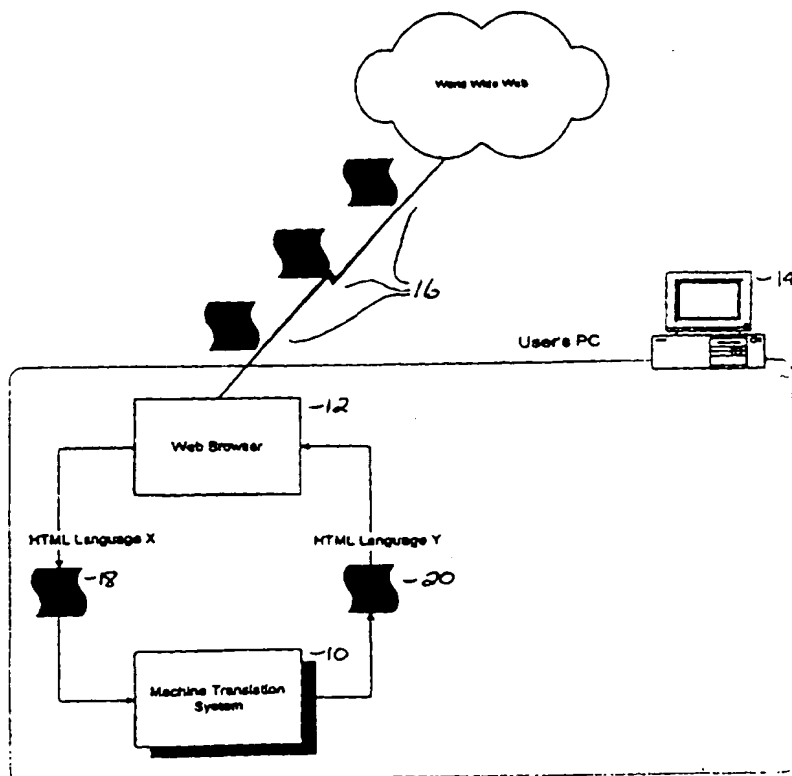
INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : G06F 17/28		A1	(11) International Publication Number: WO 97/18516
			(43) International Publication Date: 22 May 1997 (22.05.97)
(21) International Application Number: PCT/US96/18102		(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).	
(22) International Filing Date: 13 November 1996 (13.11.96)		Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	
(30) Priority Data: 08/555,916 13 November 1995 (13.11.95) US			
(71) Applicant: COMPUSERVE INCORPORATED [US/US]; 500 Arlington Centre Boulevard, Columbus, OH 43220 (US).			
(72) Inventors: FLANAGAN, Mary, A.; 558 Belknap Road, Framingham, MA 01701 (US); LAMMERS, John, A.; Apartment A3, 23 Davis Road, Acton, MA 01720 (US).			
(74) Agents: STANDLEY, Jeffrey, S. et al.; Standley & Gilcrest, Suite 500, 555 Metro Place North, Dublin, OH 43017 (US).			

(54) Title: INTEGRATED MULTILINGUAL BROWSER

(57) Abstract

The disclosed system translates into different languages HTML documents (16) available through the World Wide Web. HTML documents (16) are translated by machine translation software (10) bundled in a browser (12). Alternatively, documents are retrieved as needed, translated, and stored on a Web server so user requests are serviced with a document that has been translated from a different language. The disclosed invention expands usage of the Internet for non-English speakers.



INTEGRATED MULTILINGUAL BROWSER

5

BACKGROUND AND SUMMARY OF THE INVENTION

The present invention relates generally to the field of electronic communication over a computer network. Particularly, the present invention relates to the expansion of multi-lingual electronic communication through translation services for documents and messages available
10 through the Internet.

The recent surge in media attention to the Internet, and especially the World Wide Web, coupled with the continuing growth in home PC ownership have resulted in a growing diversity of the Internet user population. No longer is the Internet the province of software experts; thousands of novice users have begun to come online each day. Software like
15 CompuServe's Web Browser lets users quickly connect to and find useful content online. This phenomenon is not restricted to the United States or to English-speaking countries. Growth in online usage in Europe and Asia is increasing even more quickly than in the U.S.

While interest in the online world is at a peak, a significant obstacle exists to broad usage of the Internet for non-English speakers. The vast majority of Internet content is in
20 English, and is therefore inaccessible to users with other native languages. Translation of Internet documents by a human translator is not a practical solution for two reasons. First, human translation is costly and slow. A translator can typically produce 300-400 words per hour at costs of 12¢ per word or more. Second, in order to have a translator convert Internet documents to the user's native language, the user would have to download every document he
25 was interested in to provide it to the translator. This is a time-consuming process, and if the user knows no English, he will not even be able to assess the relevance of the document before

Figure 9 is a diagrammatic view of one embodiment of the present invention in which pre-translated Web pages are accessible from a server.

DETAIL DESCRIPTION OF PREFERRED EMBODIMENT(S)

5 Although the detailed description of a preferred embodiment focuses on automatic translation of World Wide Web pages, the concept is adaptable to documents obtained from other sources.

 The World Wide Web (WWW or the Web) is a distributed information system that may be accessed through a number of sources. It is comprised of software and a set of
10 protocols and conventions. Information on the Web may be accessed using a browser program such as CompuServe's Web Browser. Browsers allow users to read documents and to locate documents from other sources. They present an interface for interacting with the system and they process requests on behalf of the user.

 Information providers on the WWW make their information available through
15 programs that understand the HyperText Transfer Protocol (HTTP). Browsers assist users in "visiting" Web sites where information is stored. Information is displayed in pages of text and graphics called "Web Pages." An example of a Web page as viewed through CompuServe's Web Browser is provided in Figures 1A and 1B. The Web page shown in Figures 1A and 1B contains both text 14, 18 and graphics 10, 12, 16. The title bar 20, menu options 22, buttons
20 24, and document information 26 appearing at the top of the screen are part of the browser used to view the Web page.

 In most cases, information providers make information available through a Web server. The server responds to information requests by delivering the requested information to the user's browser for viewing. Some providers may make their information available through a

Figure 2 is the hypertext document that describes the Web page shown in Figures 1A and 1B. Figure 2 shows the markups and related words (that comprise codes) as well as data characters that may appear in a hypertext document. For example, the characters "<" and ">" appearing throughout the document are markups. The characters "<" and ">" combined with the word "head" ("`<head>`") 10 may be considered a code. Finally, the text "NLT Home" 10 that is not surrounded by markups or codes may be considered data characters.

As indicated by the brief description, HTML documents have a well-defined and documented structure defined by a grammar. The codes in a HTML document convey important information regarding both the display or presentation of the document itself as well as related references and commands. Display and presentation information may include color information, information about graphics that appear on the page, information about text that appears on the page, etc. A HTML document is structured as a series of elements that are identified by the language markups and codes. A document includes a head (consisting of a title and other optional elements) and a body that is a text flow of paragraphs, lists, images, and other elements. The various parts of the document may be identified by looking at the markups or codes in the document. For example, referring again to Figure 2 which shows the hypertext for Figures 1A and 1B, the document head contains the title "NLT Home" 10. An image contained in the document is identified in the line

```
"<br></center>" 12.
```

As may be apparent, the process of translating a HTML document requires examination of each character in document. Characters may be examined individually and in combination to determine whether they are markups, codes, or data characters. To process a document, the processing software examines the character stream that comprises the

Figure 2. In this example, the boundary markers used to identify the HTML codes are the character pairs "<." and ">.". Any character or character combination that does not normally occur in text may be used as a boundary marker. The line that appeared as "<head><title>NLT Home<title><head>" in Figure 2 (10) is preprocessed in Step 1 to the line
5 "{.<head>}.{.<title>}NLT Home{.<title>}.{.<head>}" in Figure 3 (10). Other lines in the document are preprocessed similarly.

Step 2. Machine Translation (MT) software performs the translation of text from one language to another language. There are many commercially available MT software packages. Figure 4 is an illustration of a system in which MT software 10 takes as input text in one
10 language 12 and generates a rough draft translation of the text in another language 14 using an electronic dictionary 16 and a set of linguistic and/or statistical rules encoded in the program 18. MT software can perform language conversion operations very quickly; in some cases, at speeds of up to 3,000 words per minute. The translated texts are not high quality translations, but they are usually adequate for understanding what the document is about.

15 Referring to Figures 5A and 5B, an example of a translated HTML document is shown. The HTML document of Figures 5A and 5B is the translated version of the preprocessed HTML document shown in Figure 3. As described above, the boundary markers used to identify the HTML codes are the character pairs "<." and ">.". Consequently, the MT software ignores all text that falls between the boundary markers. Data characters that are not
20 surrounded by boundary markers are translated by the MT software. The preprocessed line that appeared as "{.<head>}.{.<title>}NLT Home{.<title>}.{.<head>}" in Figure 3 (10) is translated in Step 2 to the line "{.<head>}.{.<title>}NLT Maison{.<title>}.{.<head>}" in Figure 5A (10).

that delivery time to the user may be reduced. Although storing documents requires disk space, it may represent a better use of system resources because documents that are accessed frequently are translated once rather than every time they are accessed.

Figure 9 is a diagrammatic view of an alternative implementation in which pre-translated Web pages are stored on a Web server 14. The translation software resides on a translation server 14 (possibly the same machine as the Web server). Popular Web pages 24 are pre-translated and stored in a cache 28, with additional pages being added as they are requested by users 20. The cache is a dynamic storage device with a finite capacity. New, pretranslated pages are added to the cache, but pages will also be removed from the cache if they are used infrequently or if there are constraints on storage capacity.

In accessing the system, a user 10, sends to the Web Server 14 a request for a specific page in a specific language 12. The Web Server 14 then sends a request to get the desired page 16. The method for servicing the request depends on where the page is located. If the page has been pre-translated 24 and stored in the cache of pages in multiple languages 28, it is retrieved from the cache 26 and returned to the user in the requested language 30. If the page has not been pre-translated, then the page is retrieved 20 from the World Wide Web 22, translated into the requested language, and cached before being sent to the user 30.

Translation of Web pages, in either the bundled browser/MT configuration or the Web Server configuration, requires processing of HTML codes containing reference, command, and display information. Preferably, the HTML codes are identified prior to translation, then surrounded by special boundary markers to block the translation process on the codes. The HTML preprocessor uses its knowledge regarding the markups, codes, data characters and the structure of HTML documents to determine which codes should be blocked from the translation process. After translation is complete, a postprocessing program removes the

WHAT IS CLAIMED IS:

1. A method for translating a document, comprising the steps of:
 - providing a character stream including codes and data characters in a first language;
 - 5 transmitting said character stream to a language translator;
 - recognizing said codes to prevent translation of said codes by said language translator; and
 - translating a significant portion of said data characters into a second language using said language translator.
- 10 2. The method of claim 1, wherein said codes are HyperText Markup Language codes.
3. The method of claim 1, wherein said step of recognizing said codes is performed by a language translator preprocessor.
4. The method of claim 1, wherein boundary markers are placed around said codes to prevent translation of said codes by said language translator.
- 15 5. The method of claim 1, wherein said language translator is integrated into a browser program.
6. The method of claim 1, wherein said document is pretranslated.
7. The method of claim 1, further comprising the step of viewing said translated HyperText Markup Language document with a browser.
- 20 8. A document translation system, comprising:
 - a character stream containing codes and data characters in a first language;
 - a preprocessor for marking codes in said character stream;
 - a language translator for translating into a second language said data characters in said preprocessed character stream; and

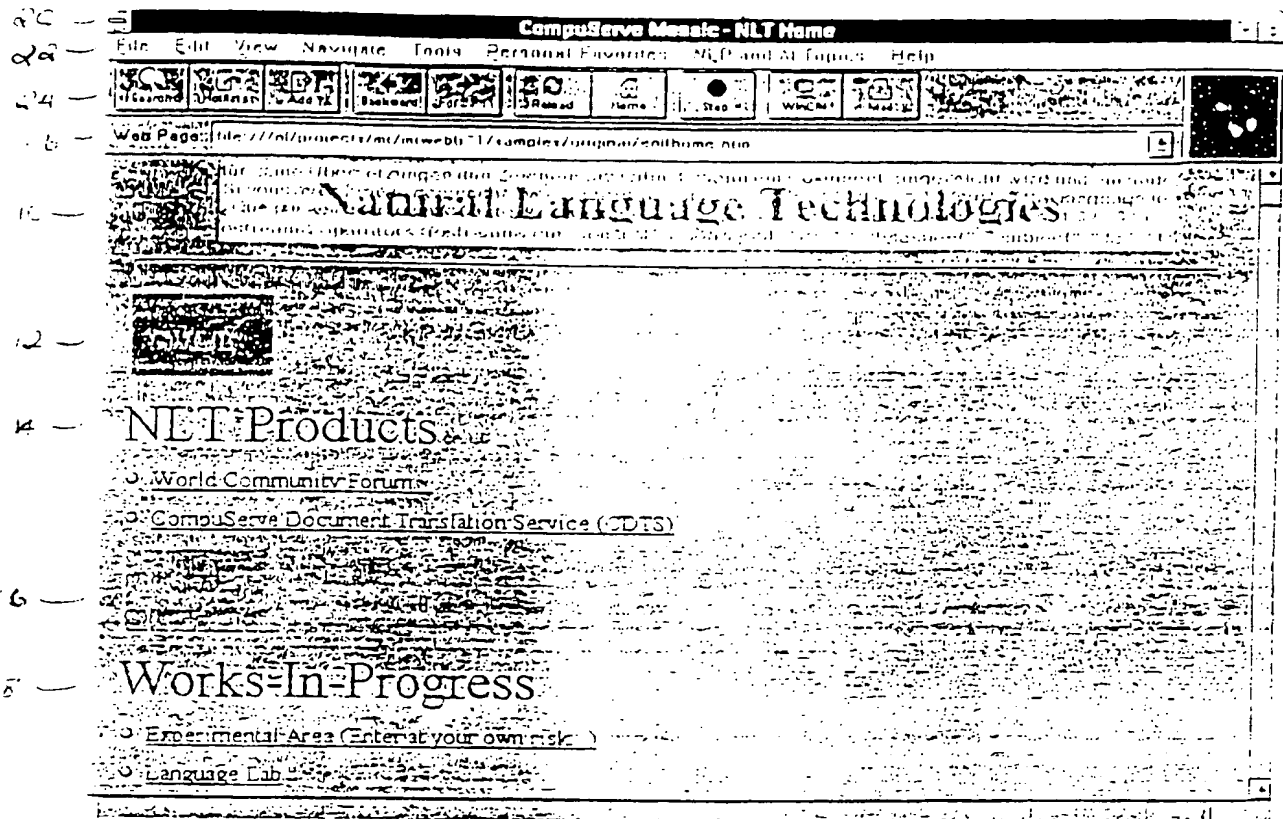


Figure 1A

```

10 - <html>
12 - <head><title>NLT Home</title></head>
    <br><center></center>
    <br>
    <center><h1> </h1></center>
    <br>
    <h1>NLT Products</h1>
    <br><UL><LI><a href="http://www.compuserve.com/cgi-bin/gocserve/CIS_EWCFORUM">World Community Forum</a></UL>
    <br><UL><LI><a href="http://www.compuserve.com/cgi-bin/gocserve/CIS_TRANSLATE">CompuServe Document Translation Service
    (CDTS)</a></UL>
    <center><h1> </h1></center>
    <br>
    <h1>Works-In-Progress</h1>
    <br><UL><LI><a href="nlttest.htm">Experimental Area (Enter at your own risk...)</a></UL>
    <br><UL><LI><a href="llab.htm">Language Lab</a></UL>
    <br><UL><LI><a href="mailtran.htm">E-Mail Translation</a></UL>
    <br><UL><LI><a href="webtrans.htm">Web Page Translation</a></UL>
    <center><h1> </h1></center>
    <br>
    <h1>Proposals</h1>
    <br><UL><LI><a href="http://jflammers/n:\projects\proposal\rose_gl.doc">Rose-Colored Glasses</a></UL>
    <center><h1> </h1></center>
    <br>
    <h1>Points of Interest</h1>
    <br><UL><LI><a href="http://www.willamette.edu:80/~jones/Language-Page.html">The Human Languages Page</a></UL>
    <br><UL><LI><a href="http://www.ai.mit.edu">MIT Artificial Intelligence Lab</a></UL>
    <center><h1> </h1></center>
    <br>
    <p>Send comments/questions to: </p>
    <br><a href="mailto:jflammers@csi.compuserve.com">NLT Mailbox</a>
    </body>
    </html>

```

Figure 2

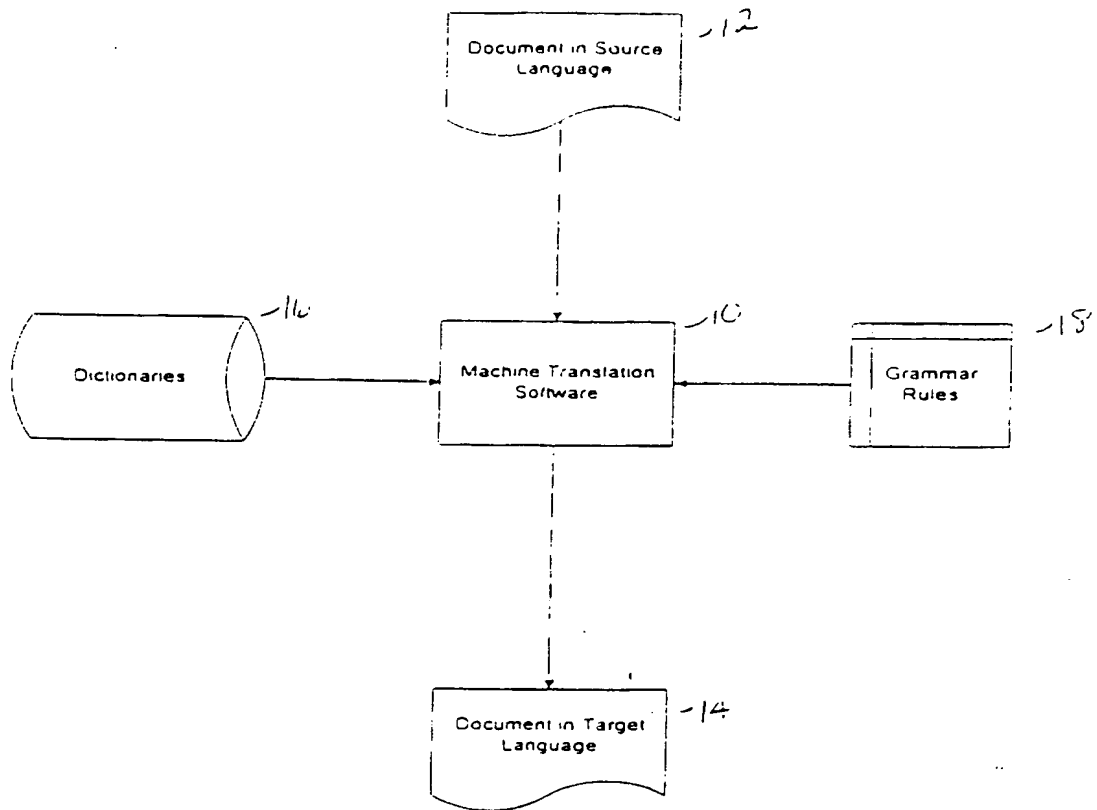


Figure 4

7/12

```

{.<center>}.{.<h1>}.{.</h1>}.{.</center> }

{.<br>}.{. }

{.<h1> } Propositions{.</h1> }

{.<br>}.{.<UL>}.{.<LI>}.{.<a
href="http://jammers/n:/projects/proposal/rose_gl.doc"> } Verres
Elevé-Colorés{.</a>}.{.</UL> }

{.<center>}.{.<h1>}.{.</h1>}.{.</center> }

{.<br>}.{. }

{.<h1> } Les points d'Intérêt{.</h1> }

{.<br>}.{.<UL>}.{.<LI>}.{.<a
href="http://www.willamette.edu:80/~tjones/Language-Page.html"> } La Page
de Langues Humaine{.</a>}.{.</UL> }

{.<br>}.{.<UL>}.{.<LI>}.{.<a href="http://www.ai.mit.edu"> } MIT le
Laboratoire d'Intelligence Artificiel{.</a>}.{.</UL> }

{.<center>}.{.<h1>}.{.</h1>}.{.</center> }

{.<br>}.{. }

{.<p> } Les commentaires/questions d'envoi à: {.</p> }

{.<br>}.{.<a href="mailto:jammers@csi.compuServe.com"> } NLT Boite à
lettres{.</a> }

{.</body> }

{.</html> }

```

Figure 5B

9/12



Figure 7A

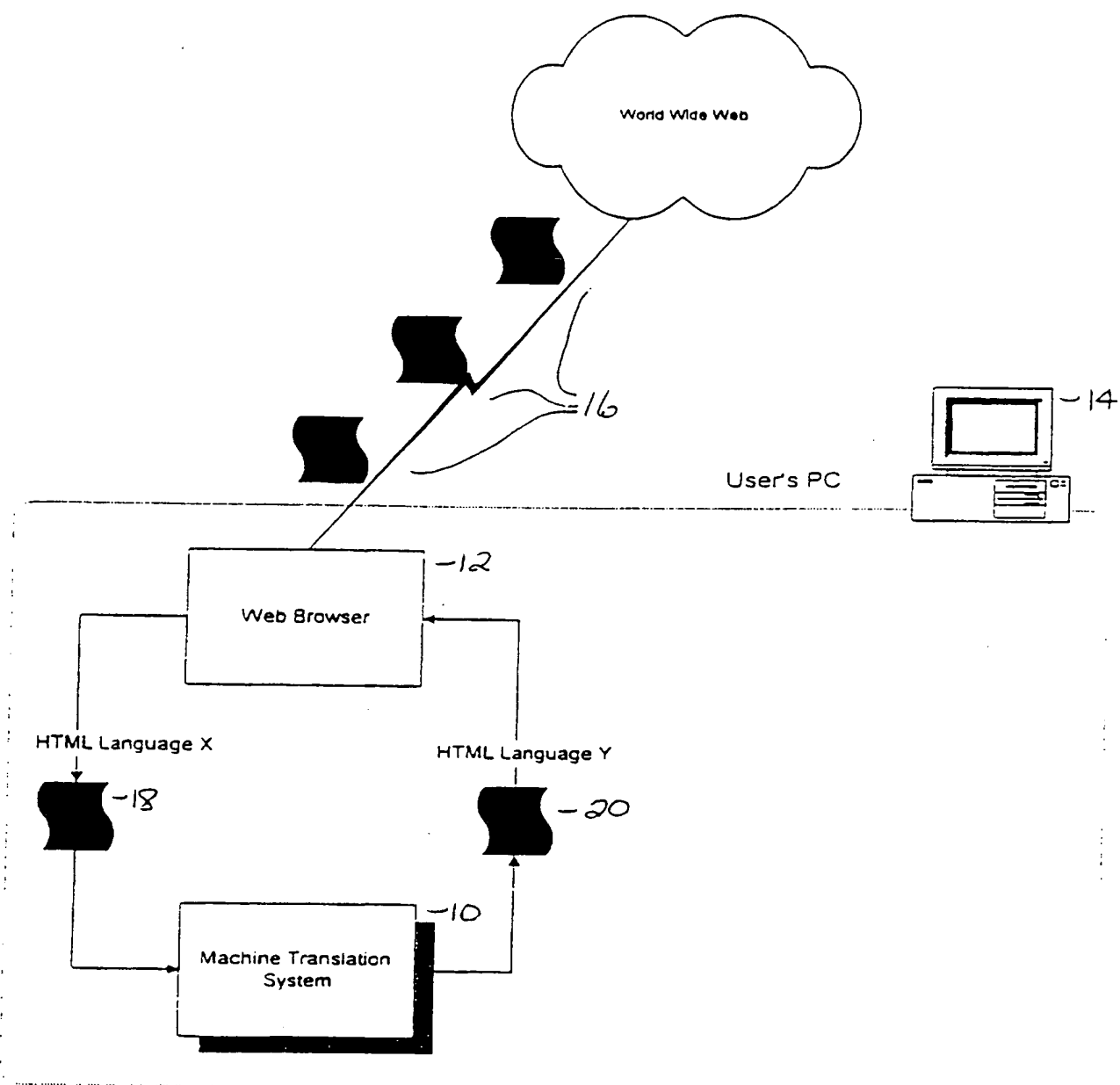


Figure 8

INTERNATIONAL SEARCH REPORT

International application No
PCT/US96/18102

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) G06F 17/28

US CL 395/752

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 395/752, 754, 755, 756, 762, 774, 778, 791

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
NONEElectronic data base consulted during the international search (name of data base and, where practicable, search terms used)
NONE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,E	US 5,587,902 A (KUGIMIYA) 24 December 1996, see the abstract, figs. 1, 10, 12 and 18.	1-13
X,P	US 5,548,508 A (NAGAMI) 20 August 1996, see the abstract, figs. 1-4.	1-13
A	US 5,361,205 A (NISHINO ET AL) 01 November 1994, see the abstract.	1-13
A	US 5,243,519 A (ANDREWS ET AL) 07 September 1993, see the abstract.	1-13
A	US 5,373,442 A (KUTSUMI ET AL) 13 December 1994, see the abstract, fig. 4B.	1-13
A	US 5,005,127 A (KUGIMIYA ET AL) 02 April 1991, see the abstract, figs. 12-13.	1-13

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	*T later document published after the international filing date or priority date and not in conflict with the application but cited to underlain the principle or theory underlying the invention
*A document defining the general state of the art which is not considered to be of particular relevance	*X document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*E earlier document published on or after the international filing date	*Y document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is considered with one or more other such documents, such combination being obvious to a person skilled in the art
*L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*Z document member of the same patent family
*O document referring to an oral disclosure, use, exhibition or other means	
*P document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

07 FEBRUARY 1997

Date of mailing of the international search report

14 MAR 1997

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

ROBERT A. WEINHARDT

Telephone No. (703) 305-3800